

Comments on Terrestrial ECOFRAM Draft Report

Michael Lavine

Accounting for Uncertainty Although this document does make some strides in accounting for uncertainty in risk assessment, I believe it doesn't go far enough. There are two main points I would like to make.

1. First, I advocate increased use of subjective probability distributions. The report makes statements such as these:

...the basic toxicity studies ...generate a single LD50 or EC50, a single dose-response slope, a single dose-response equation, and a single dose-response curve. (page 1-30)

...because the dose-response slope database is currently inadequate for extrapolation purposes, it will be assumed that the dose-response slope for the hypothetical bird species is equivalent to that of the experimental bird species. (page 2-21)

In my view, these statements don't adequately account for uncertainty. In the first case, we should acknowledge that a single toxicity study, or even a collection of toxicity studies, does not tell us precisely what the LD50 is, or the dose-response slope, etc. Toxicity studies can only tell us what values of the LD50 (or EC50 etc.) are best supported by the data, but no matter how much data are collected, we cannot know the LD50 exactly. A good way to account for the remaining uncertainty is through a probability distribution that gives the most weight to LD50's that are well supported, less weight to LD50's that are only moderately supported, and the least weight to LD50's that are poorly supported, etc. In the second case, we should acknowledge that the dose-response slope for the hypothetical species is likely to be different from the slope for the experimental species. And again, a good way to account for the uncertainty of extrapolation is through a probability distribution.

These examples are similar in spirit to other places in the report that advocate using a probability distribution only if enough data exist to provide a good estimate of the distribution, and to use a point estimate

otherwise. In other words, the report says that when we have little information we should act as though we know exactly what the value of the variable is. This strikes me as backwards. It seems better that when we have little information we should act as though we *don't* know what the value of the variable is. In fact, we shouldn't even act as though we know what the distribution is. Rather, we should entertain a range of distributions, chosen with an eye to either plausibility or conservatism, and either average over them (a Bayesian approach) or see whether the choice of distribution matters (a sensitivity analysis).

Another example is on page 3-13, which says "PT, TFIR and AV now require only a single estimate for each day" Why just a single estimate; why not a distribution?

2. Second, I advocate that estimates of dose, toxicity, risk, etc. be accompanied by estimates of their accuracy. For example, where the first paragraph of Section 1.4.2 **Analysis of Exposure and Effects (Toxicity)** says "This phase provides an exposure characterization, which includes estimates of dose and/or dose distributions ..." I would also want to see an assessment of how accurate those estimates are likely to be. Where page 1-23 says "we will define probabilistic assessments as those that estimate the ...probability that the percentage of non-target organisms adversely affected by pesticides will be (1) less than or equal to or (2) greater than any given percentage of concern" I would add that good risk assessments, probabilistic or not, should include estimates of their accuracy and give ranges of possible risks, not just point estimates.

One way to account for parameter uncertainty in predicting responses in regression models is to use prediction intervals. Another way is to use a Bayesian analysis. A discussion of the two possibilities could be added to Section 4.4.3.2 following the paragraph which now reads "... this section treats the parameters of the dose-response model ... as if they are known. ...In actuality, these parameters are subject to a range of uncertainties.

Technical Matters

1. The report is dense with references to the LD50 and slope (or equivalent parameters), as though all dose-response curves belong to a two parameter family. They don't. The report never suggests that researchers check whether their toxicity data is well fit by a probit, logit or whatever model they are using. They should.

A related point is that estimates of the LD05 can be very sensitive to whether the dose-response curve is modelled by a probit, logit, or some other family of curves. That is, it may happen that probit and logit curves are very similar in the range of doses where data were collected, so they both fit the data equally well; but they may give quite different extrapolations to the LD05. It is not sufficient to fit just one model and use it for extrapolation; a good risk assessment should account for parameter uncertainty, model uncertainty and the uncertainty of extrapolation.

2. Page 7-12 says "The avian acute oral test is well designed for producing an LD50 ... However, it may be important, ..., to develop better estimates of low levels of mortality, e.g., the LD5 or LD10." I second this point. Unfortunately, I am not aware of any statistical research on good experimental designs for estimating the LD5 or LD10. Here is a place where more statistical research would be useful.
3. Page 1-31 gives Method A of of generating a risk PDF. There is a single dose PDF and several sensitivity PDF's. It sounds as though the procedure is:

```

repeat N times: {
  repeat n times: {
    select a dose from the dose PDF
    select a sensitivity PDF
    select a sensitivity from the sensitivity PDF
    compare the dose and the sensitivity
  }
}
```

Either the procedure is wrong or I have misunderstood the intent. As I have written it, the variability among the N different runs is just the variability of the binomial distribution. The procedure should be:

```

repeat N times: {
```

```

select a sensitivity PDF
repeat n times: {
  select a sensitivity from the sensitivity PDF
  select a dose from the dose PDF
  compare the dose and the sensitivity
}
},

```

as stated on page 4-42. The difference is whether the n sensitivities in a run are all drawn from the same sensitivity PDF. If the different sensitivity PDF's represent possible PDF's that might be realized in the field then, in fact, all the true sensitivities will be drawn from just one of them, and the Monte Carlo simulation should do the same. Is this an important distinction? Yes. It's easy to construct examples where the distinction is important.

4. Page 1-26 suggests that Monte Carlo simulations account for "correlations between any of the input variables" through a correlation matrix. This is true as far as it goes, but fails to account for nonlinear dependence between variables which, as mentioned on page 1-16, may be important.
5. The report is full of references to Monte Carlo simulations. It should be recognized that Monte Carlo simulation is just one of many ways to perform a mathematical integration. The omnipresence of Monte Carlo simulation in the report may give the impression that it is either the preferred method, or perhaps the only method that the EPA accepts. But for many integrals Monte Carlo simulation is not the best method. The report could explain how Monte Carlo simulation is equivalent to evaluating an integral and make clear that other methods are acceptable.
6. Page 3-13 says "these six parameters are averaged over all fields ...". Averaging will work well if the result is used linearly. (I couldn't tell whether it is used linearly.) But in general, if g is a nonlinear function, then the average of $g(x)$ (averaged over values of x) is not equal to $g(\bar{x})$. Page 3-25 says "One option is therefore to substitute $FMR/(GE \times AE)$ for TFIR. Here is an example where point estimates are not used linearly. That is, $TFIR = FMR/(GE \times AE)$ is a nonlinear function

of both GE and AE, so substituting in average or typical values of GE and AE might not yield an average or typical value of TFIR. One can't tell in advance whether this is an important problem in any given risk assessment. One simply has to look at the values and uncertainties in GE and AE and make a sensitivity analysis. An alternative is to keep GE and AE as variables to be randomly selected in the Monte Carlo analysis.

7. Section 4.4.2.1 uses the ratio of standard error to slope as a measure of within-test variability and says "use of the reported error on the estimate from any given test would account for most of the variability expected across tests." I'm not sure what's meant by this, but consider what would happen if tests were run with more animals. Standard errors would go down to 0 if enough animals were used, but variability across tests would not go to 0. So I don't see how use of the reported error could correctly account for cross-test variability in general.
8. Section 4.4.2.1 discusses sources of variability for toxicity measurements. This seems like a modelling problem well suited to hierarchical models. An example would be something like this:
 - (a) within a lab there is variability among different tests of the same chemical on the same species. Let's say, for a simple example, that the LD50's for the i 'th species from tests done by the j 'th lab are Normally distributed about some unknown mean $\mu_{i,j}$, and with variance σ_l^2 ;
 - (b) the $\mu_{i,j}$'s vary between labs, having mean μ_i and variance σ_i^2 ;
 - (c) the μ_i 's vary among species, having mean μ and variance σ_s^2 .

Similar hierarchical models have become popular over the last decade as new computational techniques have made them easier to fit, both classically and Bayesianly. They are appealing because they model explicitly how information from one (species, lab, test) combination gives us information about another (species, lab, test) combination.

Further Questions

1. Page 2-11 suggests consulting the label for each product to establish potential exposure issues. Are products always applied according to the label?
2. Page 2-13 foretells a database of model environments. I wonder whether this is overly optimistic. Is there precedent?
3. The bottom of page 3-7 talks about a “cell’ model of habitat structure.” I didn’t see references to any other models or any other ways of doing risk assessment. Is a cell model the only way? Is it mandated? I’m concerned because sometimes an overly mechanistic way of modelling can lead to inferior results, if too little is known about the mechanism. As an example, if the distribution of dose can be estimated directly from field studies, such an estimate might be better than one coming from a cell model.